

## 工作周报 2012.12.02 - 2012.12.08

徐佳逸

### 本周小结

任务是熟悉 Uncertainty 项目，对代码做一些更新和修改，看下数据规律。

### 科研进展

1. 熟悉海东师兄以前写的代码；
2. 更改应用的计算两个分布之间距离的公式，从

$$d(P, Q) = \lambda E(\bar{P}, \bar{Q}) + \theta J(P||Q)$$

更改为

$$d(P, Q) = \frac{1}{2} (J(P - \bar{P} || Q - \bar{Q}) + J(\bar{P} || \bar{Q}))$$

$$0 \leq d(P, Q) \leq 1$$

3. 更新对应的计算代码；
4. 生成了两组数据

数据参数都为(2, 10, 100, 3):

两维、10 个不确定的对象、每个对象采集 100 个样本，分属 3 个聚类

这 3 个聚类的统计数据为：(6, 0.5), (7, 0.5), (8, 0.5)，两个数值分别代表均值和方差

用 1 代表正态分布，2 代表高斯分布，3 代表逆高斯分布

- (1) 数据 1 的 10 个对象的分布类型分别为：

(3, 2, 1, 1, 3, 1, 3, 1, 3, 3)

- (2) 数据 1 的 10 个对象的分布类型分别为：

(1, 2, 3, 2, 3, 2, 3, 2, 3, 3)

5. 规律：

- (1) 正态分布：

当两个分布参数相同时，距离一般能控制在小于 0.05，但两个不同均值例如 6 和 7，两个正态分布的距离也可能很小，一般在 0.05-0.1 之间；

正态与高斯分布的距离很近，当均值和方差的参数一样时，距离可以达到 0.4 以下；

正态分布与逆高斯分布的差异明显；

- (2) 高斯分布：

当两个分布参数相同时，距离一般能控制在小于 0.05，不同均值差异明显；

高斯分布与逆高斯分布差异明显；

(3) 逆高斯分布:

即使两个分布参数相同, 距离也只能控制在 0.4 以下 (下面会测试提升样本数是否能减少距离)

## 下周计划:

### 1. Uncertainty

- (1) 测试增加样本数后的效果;
- (2) 绘制数据集的图看看 2 维上的投影分布效果;
- (3) 测试新的数据集。

### 2. Community Detection

- (1) 从昱新那儿得到数据产生的程序, 并对数据做测试;
- (2) 找数据的规律;
- (3) 优化代码的效率。